

Input file F1h14273new; Duput File F1h14273tra
Sequence length 1743

TCCGGACTAGTTCTAGACCGCTGCGGGCGCCAGCGCCGGGA	ATG	TCC	CCT	GAA	TGC	C	A	R	A	A	GCG	27
G D A P L R S L E Q A N R T R F F S												29
GGC GAC GCG CCC TTG CGC AGC CTG GAG CAA GCC AAC CGC ACC CGC TTT CCC TTC TTC TCC												87
D V K G D H R L L V L A A V E T T V L V L												49
GAC GTC AAG GGC GAC CAC CGG CTG GTG CTG GCC GCG GTG GAG ACA ACC GTG CTG GTG CTC												147
I F A V S L L G N V C A L V L V A R R R												69
ATC TTT GCA GTG TCG CTG CTG GGC AAC GTG TGC TGC GCG GTG GCG CGC CGA CGA												207
R R G A T A C L V L N L F C A D L L F I												89
CGC CGC GGC GCG ACT GCC TGC CTG GTA CTC AAC CTC TTC TGC GCG GAC CTG CTC TTC ATC												267
S A I P L V L A V R W T E A W L L G P V												109
AGC GCT ATC CCT CTG GTG CTG GCC GTG CGC TGG ACT GAG GCC TGG CTG CTG GGC CCC GTT												327
A C H L L F Y V M T L S G S V T I L T L												129
GCC TGC CAC CTG CTC TTC TAC GTG ATG ACC CTG AGC GGC AGC GTC ACC ATC CTC ACG CTG												387
A A V S L E R M V C I V H L Q R G V R G												149
GCC GCG GTC AGC CTG GAG CGC ATG GTG TGC ATC GRG CAC CTG CAG CGC GGC GTG CGG GGT												447
P G R R A R A V L L A L I W G Y S A V A												169
CCT CGG CGG GCG GCA GTG CTG CTG GCG CTC ATC TGG GCC TAT TCG GCG GTC GCC												507
A L P L C V F F R V V P Q R L P G A D Q												189
GCT CTG CCT CTC TGC GTC TTC TTT CGA GTC CTC CCG CAA CGG CTC CCC GGC GAC CAG												567

TO FIG. 1B.

FIG. 1A.





10075907 JUL 22 2002

2/15

COPY OF PAPERS
ORIGINALLY FILED

FROM FIG. 1A.

E	I	S	I	C	T	L	L	I	W	P	T	I	P	G	E	I	S	W	D	V	209
GAA	ATT	TCG	ATT	TGC	ACA	CTG	ATT	TGG	CCC	ACC	ATT	CCT	GGA	GAG	ATC	TCG	TGG	GAT	GTC	627	
S	F	V	T	L	N	F	L	V	P	P	G	L	V	I	V	I	S	Y	S	K	229
TCT	TTT	GTT	ACT	TTG	AAC	TTC	TTG	GTG	CCA	GGA	CTG	GTC	ATT	GTG	ATC	AGT	TAC	TCC	AAA	687	
I	L	Q	I	T	K	A	S	R	K	R	L	L	T	V	S	L	A	Y	S	E	249
ATT	TTA	CAG	ATC	ACA	AAG	GCA	TCA	AGG	AAG	AGG	CTC	ACC	GTA	AGC	CTG	GCC	TAC	TCG	GAG	747	
S	H	Q	I	R	V	S	Q	Q	D	F	R	L	F	R	T	L	F	L	L	269	
AGC	CAC	CAG	ATC	CGC	GTG	TCC	CAG	CAG	GAC	TTC	CGG	CTC	TTC	CGC	ACC	CTC	TTC	CTC	CTC	807	
M	V	S	F	I	M	I	W	S	P	I	I	I	T	I	L	L	I	L	I	208	
ATG	GTC	TCC	TTC	TTC	ATC	ATG	TGG	AGC	CCC	ATC	ATC	ATC	ACC	ATC	CTC	CTC	ATC	CTG	ATC	867	
Q	N	F	K	Q	D	L	V	I	W	P	S	L	F	F	W	V	V	A	F	309	
CAG	AAC	TTC	AAG	CAA	GAC	CTG	GTC	ATC	TGG	CCG	TCC	CTC	TTC	TTC	TGG	GTG	GTG	GCC	TTC	927	

TO FIG. 1C.

FIG. 1B.



10075907 1078202

3/15

COPY OF PAPERS
ORIGINALLY FILED

FROM FIG. 1B.

T F A N S A L N P I L Y N M T L C R N E 329
ACA TTT GCT AAT TCA GCC CTA AAC CCC ATC CTC TAC AAC ATG ACA CTG TGC AGG AAT GAG 987
W K K I F C C F W F P E K G A I L T D T 349
TGG AAG AAA ATT TTT TGC TGC TTC TGG TTC CCA GAA AAG GGA GCC ATT TTA ACA GAC ACA 1047
S V K R N D L S I I S G • 362
TCT GTC AAA AGA AAT GAC TTG TCG ATT ATT TCT GGC TAA 1086

TTTTTCTTTATAGCCGAGTTTCICACACCTGGCGAGCTGTGGCATGCTTTTAAACAGAGTTCATTTCCAGTACCCCTCCA
TCAGTGCACCCTGCTTTAAGAAAAAGAACCTATGCAATAGACATCCACAGCGTCGGTAAATTAAGGGTGATCACCAA
GTTTCATAATATTTCCCTTTATAAAGGATTTGTTGGCCAGGTGCAGTGGTTCATGCTGTAAATCCAGCAGTTTGGG
AGGCTGAGGTGGTGATCACCTGAGGTGAGGAGTTCGAGACCAACCTGACCAACATGGTGAGACCCCGTCTCTACTA
AAAAATAAAAAAAAATTAGCTGGGAGTGGTGGGCACCTGTAAATCCTAGCTACTTGGGAGGCTGAACCCAGGAGAAAT
CTCTTGAACTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCGTGCCATTGCACTCCAACAGGGCAACAAGAGTGAAAC
TCCATCTTAAAAAAAAGAGATTGTTATGGGTTCCTTTTAAATGTGAACCTTTTGTAGTGTGTTGTAATATG
ATCAAAATTAATAATATTTATGACTGTTTCAGCAAAAAAAAAGGGCGG

RTA01/2057957v1

FIG. 1C:



10025987.072202

4/15

COPY OF PAPERS
ORIGINALLY FILED

Sequence	Description	score	E-value	N
7tm_1	PFO0001 7 transmembrane receptor (rhodopsin	119.9	4.7e-37	1

Parsed for domains:

Sequence	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value	-
7tm_1	1/1	57	321	1	259 []	119.9	4.7e-37	

Alignments of top-scoring domains:

7tm_1: domain 1 of 1, from 57 to 321: score 119.9, E = 4.7e-37

*->GN1LVilvilrtkk1rtptnifi1NLAVADLLf11t1ppwalyylvg
GN+ +++++r +tr +t +t+1NL ADLLf + p++ ++ -+

F1h14273, 57 GNVCALVLVAR-RRRRGATACLVLNLF CADLLFISAIPVLAVR-WT 101

gaadWpfGsa1Ck1vtaldvvnmyaSi11Lta1SiDRY1A1vhP1ryrrr
e W++G++ C+1+ ++++++ + i1+L+a S++R + Iv 1+ +r

F1h14373, 102 --EAWLLGPVACHLLFYMTLSGSVTILTAAVSLERMVCIV-HLQRGVR 148

rtsprrAkvv11vwv1a111s1Pp11fswvktveegngt1nnvnevC1i
+r +v+++1+W +++++1P +f+ v+ ++ ++ ++ +C++

F1h14273, 149 GPGRRARAVLLALIWGYSAAALPLCVFFRVVPQRLPG--ADQEISICTL 196

dfpccstasvstwlrsvv11st1wgF11P11vilvcYtrI1rt1r....
+p++++ ++st +++ ++ F1+P 1vi++ Y+ I1 + + +++

F1h14273, 197 IWPTIPG---EISWVVSFVTLNELVRGLVIVISYSKILQITKasrkr 240

.....kaakt11vvvvvFv1CW1Pyfiv111dt1c
+ + +++++ + +++++ ++ +t1++++v F++ W P i++11 +

F1h14273, 241 1cvslayseehqirvsqqdfrLFRTLFLLMVSFFIMWSPIIITILLILIQ 290

.1siimsstCelervipta11vt1wLayvNsc1NPi1Y<-*
-+ + + p +++++ + +++++Na+1NPi+Y

F1h14273, 291 nFK-----QDLVIWPSLFFWVAPTAFANSALNPILY 321

FIG. 2.



5/15

COPY OF PAPERS
ORIGINALLY FILED

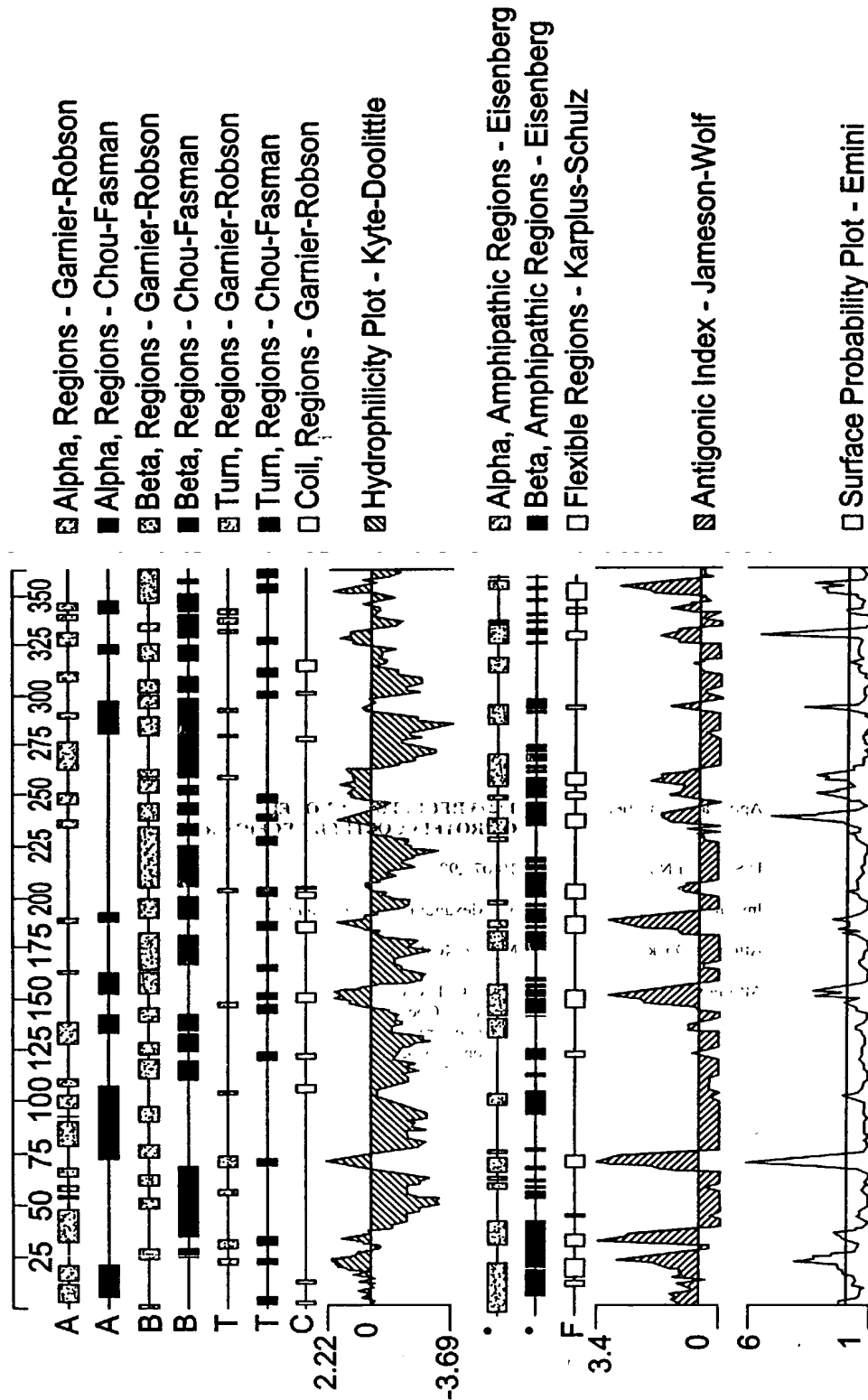


FIG. 3.

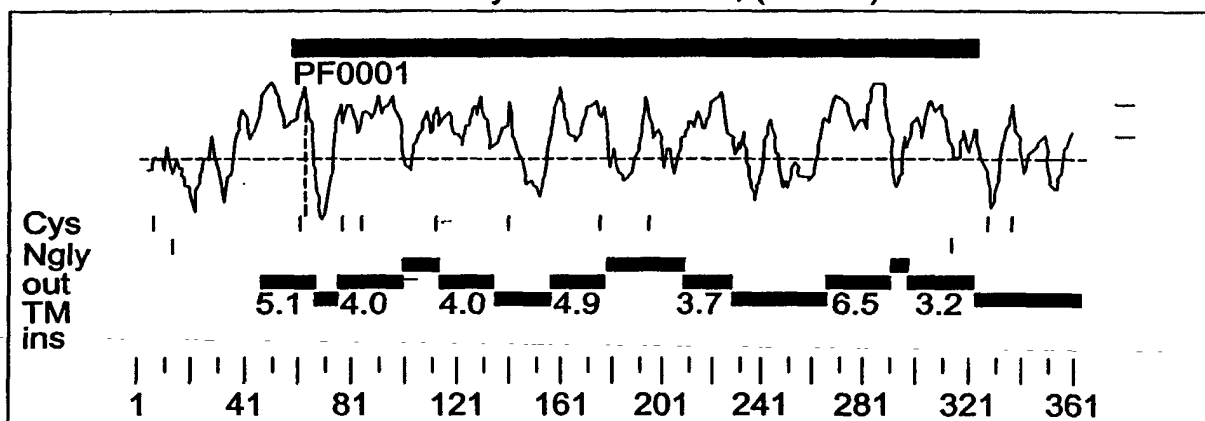


30075407 1072202

6/15

COPY OF PAPERS
ORIGINALLY FILED

Analysis of Flh14273, (362 aa)



>F1h14273, 1086 bases, 1825 checksum.
MSPECARAAQDAPLRSLEQANRTRFPFFSDVKGDHRLVLA AVETT VLVLEFAVSL LGNVC
ALVLVARRRRRGATAQLVLNLF CADLLFYSAIPLVLAVRWTEAVLLGPVACHLLFYVMTL
SGSVTILTLAAVSLERMVCIVHLQRCVRGPGRRARAVLLALIWGYSAAALPLCVFFRVV
PQRLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPGLVIVISYSKILQITKASRR
LTVSLAYSESHQIRVSQQDFRLFRTEFLLMVSFFIMWSPITITILLILIQNFQDLVIWP
SLFFWVVAFTFANSALNPILYNMTCRNEWKKIFCCPWFGEKGAILTDTSVKRNDLSIIS
G+

FIG. 4.



00075967.072P02

COPY OF PAPERS
ORIGINALLY FILED

7/15

Prositate Pattern Matches for F1h14273

>PS000001/PDDC000001/ASN_GLYCOSYLATION N-glycosylation site.

Query: 21 NRTR 24

Query: 322 NMTL 325

>PS000004/PDDC000004/CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylatic

Query: 239 KRLT 242

>PS000005/PDDC000005/PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 237 SRK 239

Query: 350 SVK 352

>PS000006/PDDC000006/EK2_PHOSPHO_SITE Casien kinase II phosphorylation site.

Query: 256 SQQD 259

>PS000008/PDDC000008/MYRISTYL N-myristoylation site.

Query: 57 GNVCL 62

Query: 72 GATACL 77

Query: 343 GAILTD 348

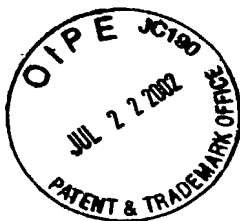
>PS000009/PDDC000009/AMIDATION Amidation site.

Query: 150 PCRR 153

>PS000029/PDDC000029/LEUCINE_ZIPPER Leucine zipper pattern.

Query: 106 LGPVACHLLFYVMTLSGSVTIL 127

FIG. 5.



100759132 10722102

COPY OF PAPERS
ORIGINALLY FILED

8/15

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
75	98	ins->out	4.0
113	134	out->ins	4.0
156	177	ins->out	4.9
209	227	out->ins	3.7
266	289	ins->out	6.5
297	321	out->ins	3.2

>F1h14273,

MSPECARAAGDAPLRSLERANRTRFPFFSDVKGDHRLVLA AVETTVLVLIPAVSLLGNVC
ALVLVARRRRRGATACLVNLF CADLLFISAWPLVLAVRWTEAWLLGPVACHLLFYVMTL
SGSVTILTLAAVSLERMVCIVHLQRGVRGPGRRARAVLLALIWGYSAAALPLCVFFRVV
PQRLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPGLVIVISYSKILQITKASRKR
LTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDLVIWP
SLFFWVVAFTFANSALNPILYNMTLCRNEWKKIFCCFWFPEKGAILTDTSVKRNDLSIIS
G

Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
14	37	ins->out	4.0
52	73	out->ins	4.0
95	116	ins->out	4.9
148	166	out->ins	3.7
205	228	ins->out	6.5
236	260	out->ins	3.2

>F1h14273, _mature

LVLVARRRRRGATACLVNLF CADLLFISAIPLVLAVRWTEAWLLGPVACHLLPYVMTLS
GSVTILTLAAVSLERMVCIVHLQRGVRGPGRRARAVLLALIWGYSAAALPLCVFFRVVP
QRLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPGLVIVISYSKILQITKASRKR
LTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDLVIWPS
LFFWVVAFTFANSALNPILYNMTLCRNEWKKIPCCFWFPEKGAILTDTSVKRNDLSIISG

FIG. 6.



10075917-07P102

COPY OF PAPERS
ORIGINALLY FILED

9/15

Input file 14273mj Output File 14273mtra
Sequence length 1560

```
TTGCCAAGCTCAGCGTAAGCCTCTTCCACTGCAATCTCACAGAGGGGTTTCATGGAGTGCTTCACACCATCAGTGACCA
CTCCAGACTTGTCGGGCTTACCCGAATCTTCACAGCGGAGTCGATGACCCCTCTTGACAGCCACGAGCGCGCAGCTC
M S P E C A Q T T G 10
CGCCATCTCCCGGACGCGTGGCGCGGCGCGGC ATG TCC CCT GAG TGT GCA CAG ACG ACG GGC 30
P G P S H T L D Q V N R T H F P F S D 30
CCT GCT CCC TCG CAC ACC CTG GAC CAA GTC AAT CGC ACC CAC TTC CCT TTC TTC TCG GAT 90
V K G D H R L V E S V V E T T V L G L I 50
GTC AAG GGC GAC CAC CGG TTG GTG TTG AGC GTC GTG GAG ACC ACC GTT CTG GGA CTC ATC 150
F V V S L L G N V C A L V L V A R R R 70
TTT GTC GTC TCA CTG CTG GGC AAC TGG TGT GCT CTA GTG CTG GTG GCG CGC CGT CGG CGC 210
R G A S A S L V L N L F C A D L L F T S 90
CGT GGG GCG TCA GCC AGC CTG GTG CTC AAC CTC TTC TGC GCG GAT TTG CTC TTC ACC AGC 270
A I P L V L V V R W T E A W L L G P V V 110
GCC ATC CCT CTA GTG CTC GTG GGC TGG ACT GAG GCC TGG CTG TTG GGG CCC GTC GTC 330
C H L L F Y V M T M S G S V T I L T L A 130
TGC CAC CTG CTC TTC TAC ATG ACA ATG AGC AGC AGC GTC ACG ATC CTC ACA CTG GCC 390
A V S L E R M V C I V R L R R G L S G P 150
GCG GTC AGC CTG GAG CGC ATG GTG TGC ATC GTG CGC CTC CGG CGC GGC TTG AGC GGC CCG 450
G R R T Q A A L L A F I W G Y S A L A A 170
GGG CGG CGG ACT CAG GCG GCA CTG CTG GCT TTC ATA TGG GGT TAC TCG GCG CTC GCC GCG 510
```

TO FIG. 7B.

FIG. 7A.



1.0075997.072212

COPY OF PAPERS
ORIGINALLY FILED

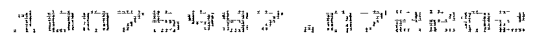
10/15

FROM FIG. 7A.

L P L Y I L F R R V V P Q R L L P G G D Q E 190
CTG CCC CTC TAC ATC TTG TTC CGC GTG GTG CCG CAG CGC CTT CCC GGC GGG GAC CAG GAA 627
I P I C T L L D W P N R I G E I S W D V F 210
ATT CCG ATT TGC ACA TTG GAT TGG CCC AAC CGC ATA GGA GAA ATC TCA TGG GAT GTG TTT 630
F E T L N F L V P G L V I V I S Y S K I 230
TTT GAG ACT TTG AAC TTC CTG GTG CCG GGA CTG GTC ATT GTG ATC AGT TAC TCC AAA ATT 690
L Q I T K A S R K R L T L S L A Y S E S 250
TTA CAG ATC ACG AAA GCA TCG CGG AAG AGG CTT ACG CTG AGC TTG GCA TAC TCT GAG AGC 750
H Q I R V S Q Q Q D Y R L F R T L L F L L M 270
CAC CAG ATC CGA GTG TCC CAA CAA GAC TAC CGA CTC TTC CGC AGC CTC TTC CTG CTC ATC 810
V S F I M W S P I I I T I L L I L I Q 290
GTT TCC TTC TTC ATC ATG TGG AGT CCG ATC ATC ACC ATC CTC CTC ATC TTG ATC CAA 870
N F R Q D L V I W P S L F F F W V A F T 310
AAC TTC CCG CAG GAC CTG GTC ATC TGG CCA TCC CTT TTC TTC TGG GTG GTG GCC TTC ACG 930
F A N S A L N P I L Y N M S L F R N E W 330
TTT GCC AAC TCT GCC CTA AAC CCC ATA CTG TAC AAC ATG TCG CTG TTC AGG AAC GAA TGG 990
R K I F C C F F F P E K G A I F T D T S 350
AGG AAG ATT TTT TGC TGC TTC TTT TTT CCA GAG AAG GGA GCC ATT TTT ACA GAT ACG TCT 1050
V R R N D L S V I S S 362
GTC AGG CGA AAT GAC TTG TCT GTT ATT TCC AGC TAA 1086

CTAGCCTCTGGTGCCAGGTGAACCACCGGTGTGTCATGTAAAGCGAGTTAACTTCAAGGAAGCCACCAGTGCCTCTGC
TTTAAAAATACCCGACTTCCAACAGCAGGCATCTACGGAGCCAGCAAAATTAAGGAATGATCGCTCAGTATAAAAATATT
TTTCCTTAAAGAACTTCTATGGTTCCTTTTGTGAACCTTTTAAAGTGTTTGTAAATATGATCTAGTTAATAAATT
TTTATTTATACGTGTTCTCTACAAAAAATAAAAAA

FIG. 7B.



**COPY OF PAPERS
ORIGINALLY FILED**

Scores for sequence family classification (score includes all domains):

Parsed for domains:

Alignments of top-scoring domains:

7tm_1: domain 1 of 1, from 57 to 321: score 118.8, E = 1e-36

14273m, 57 GNVCALVLVAR-RRRRGASASLVNLFCADLLFTSAIPLVLVVR-WT 101

14273m, 102 --EAWLLGPVVCHLLPYVMTMSGSVTILTLAAVSLERMVCIV-RLRRGLS 148

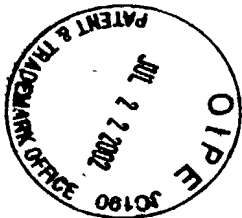
14273m, 149 GP-GRRTqAALLAFIWGYSALAALPLYILFRVVPQRLPGGD--QEITPCT 195

14273m, 196 LDWPNRIG-----EISWDVFFETLNFLVPGLVIVISYSKILQITKasrk 239

14273m,240 rltt1slayseshqirvsqdyRLFRTLFLLMVSFFIMWSPIIITILLILI 289

14273m,290 QnFR-----QDLVIWPSLFFWVVAFTFANSALNPILY 321

FIG. 8.



10075987 072202

COPY OF PAPERS
ORIGINALLY FILED

12/15

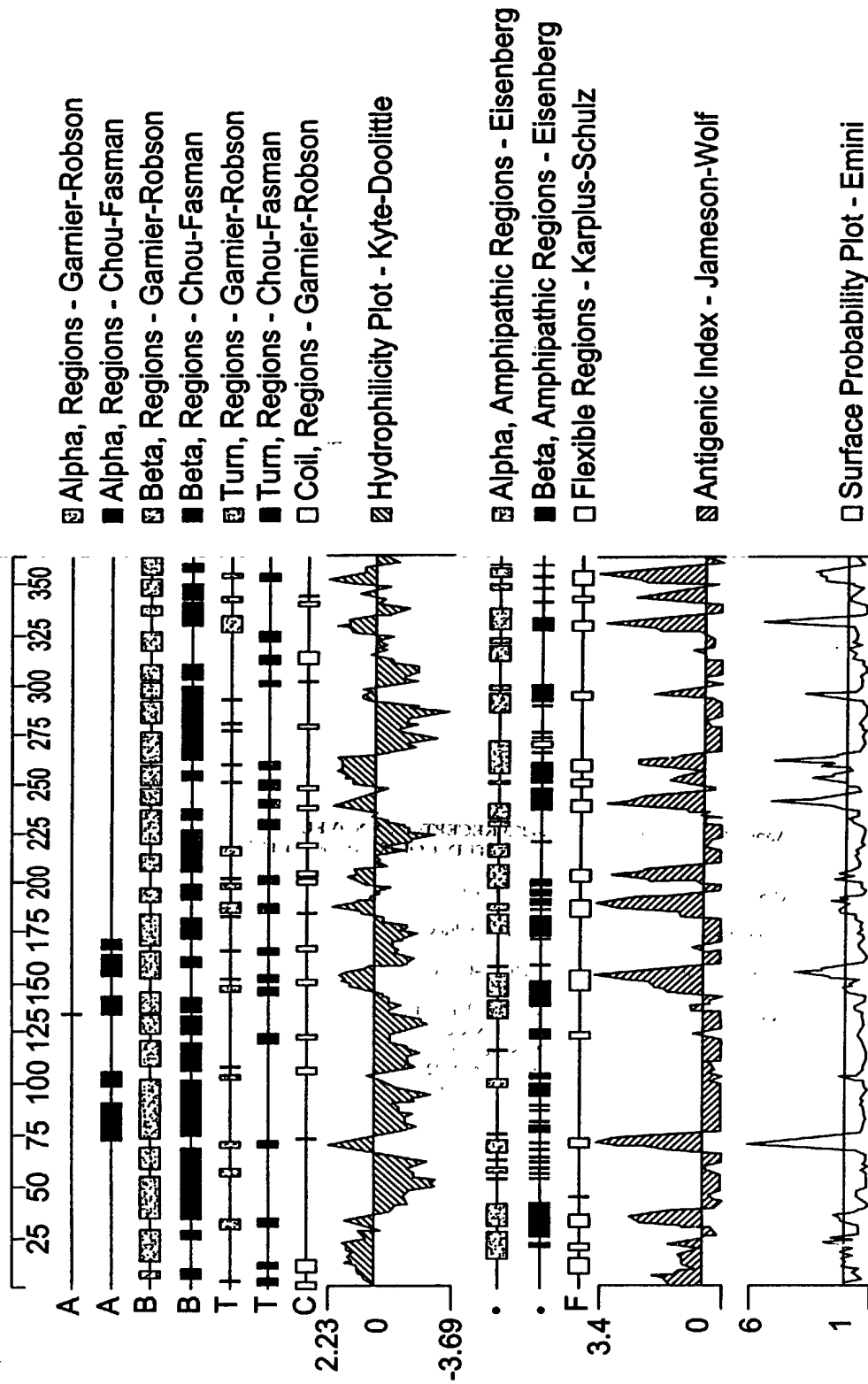


FIG. 9.

50 75 100 125 150 175 200 225 250 275 300 325 350

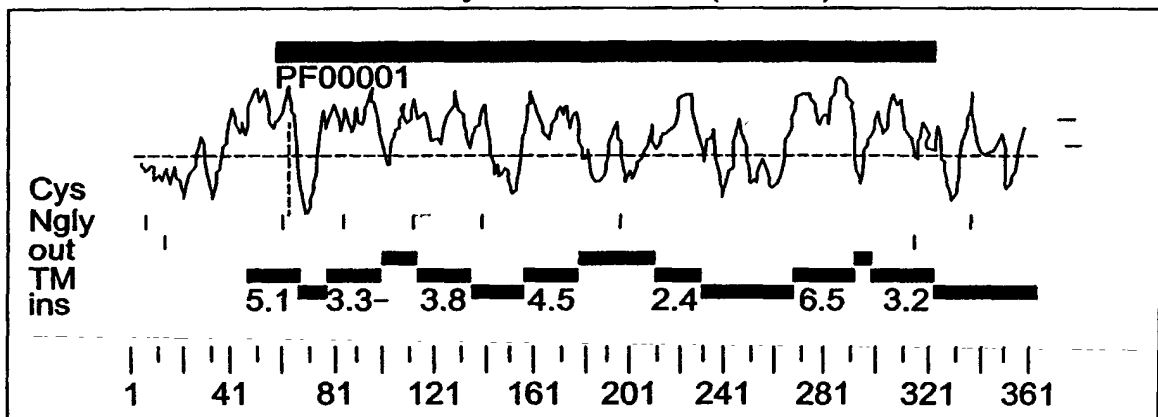


10075987.072F02

13/15

COPY OF PAPERS
ORIGINALLY FILED

Analysis of 14273m, (362 aa)



>14273m, 1086 bases, 6943 checksum.
MSPECAQTGPCPSHTLDQVNRTHFPFFSDVKGDHRLVLSVVETTGLIFVVSLLGNVC
ALVLVARRRRRGASLVLNLF CADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFVVM TM
SGSVTILTLAAVSLERMVCIVRLRRGLSGPGRRTQAALLAFIWGYSALAALPLYILFRVV
PQRLPGGDQEIPIC TL DWP NRIGEISWDVFFETLNLV PGLVIVISYSKILQITKASRKR
LTL SLAYSESHQIRVSQQDYRLFTLFLMV SFFI WSP I I I T I L L I L I Q N F R Q D L V I W P
SLFFWVVAFTFANSALNPILYNMSLFRNEWRKIFCCFPPEKGAIFTDTSVRRNDLSVIS
S*

FIG. 10.



100025482 1072202

14/15

COPY OF PAPERS
ORIGINALLY FILED

Prosite Pattern Matches for 14273m,

>PS00001/PDDC00001; ASN_GLYCOSYLATION N-glycosylation site.

Query: 21 NRTN 24
Query: 322 NRTN 325

>PS00002/PDDC00002/GLYCOSAMINOGLYCAN Glycosaminoglycan attachment site.

RU Additional rules:
RU There must be at least two acidic amino acids (Glu or Asp) from -2 to
 -4 relative to the serine.

Query: 148 SGPG 151

>PS00004/PDDC00004/CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylation

Query: 239 KRLT 242

>PS00005/PDDC00005/PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 237 SRK 239
Query: 350 SVR 352

>PS00006/PDDC00006/CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 40 SVVE 43
Query: 256 SQQD 259

>PS00008/PDDC00008/MYRISTYL N-myristoylation site.

Query: 57 GNVCL 62
Query: 72 GASASL 77
Query: 343 GAIFTD 348

>PS00009/PDDC00009/AMIDATION Amidation site.

Query: 150 PGRR 153

FIG. 11.



100754937.072202

15/15

COPY OF PAPERS
ORIGINALLY FILED

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
77	98	ins->out	3.3
113	134	out->ins	3.8
156	177	ins->out	4.5
209	227	out->ins	2.4
266	289	ins->out	6.5
297	321	out->ins	3.2

>14273m,

MSPECAQTTGPGPSHTLDQVNRTHFPFFSDVKGDRHLVLSVVETTVLGLIFVVSLLGNVC
ALVLVARRRRRGASASLVNLFCADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFYVMTM
SGSVTILTLAAVSLERMVCIVRLRRGLSGPGRRTQAALLAFIWGYSALAALPLYILFRVV
PQRLPGGDQEIPICITLDWPNRIGESISWDVFFETLNLFLVPGLVIVISYSKILQITKASRKR
LTLSLAYSESHQIRVSQQDYRLFRTLFLLMVSFFIMWSPIIITILLILIQNFRQDLVIWP
SLFFWVVAFTFANSALNPILYMMSLFRNEWKIFCCFFPEKGAIFTDTSVRRNDLSVIS
S

Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
16	37	ins->out	3.3
52	73	out->ins	3.8
95	116	ins->out	4.5
148	166	out->ins	2.4
205	228	ins->out	6.5
236	260	out->ins	3.2

>14273m, _mature

LVLVARRRRRGASASLVNLFCADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFYVMGMS
GSVTILTLAAVSLERMVCIVRLRRGLSGPGRRTQAALIAFIWGYSALAALPLYILFRVVP
QRLPGGDQEIPICITLDWPNRIGESISWCVFFETLNLFLVPGLVIVISYSKILQITKASRKR
TLSLAYSESKQIRVSQQDYRLFRTLFLLMVSFFIMWSPIIITILLILIQNFRQDLVIWPS
LFFWVVAFTFANSALNPILYNMSLFRNEWKIFCCFFPEKGAIFTDTSVRRNDLSVISS

FIG. 12.